
On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity

Khang Le^{*, \diamond}

Huy Nguyen^{*, \dagger}

Khai Nguyen ^{\diamond}

Tung Pham ^{\dagger}

Nhat Ho ^{\diamond}

University of Texas, Austin ^{\diamond}

VinAI Research ^{\dagger}

Abstract

We study the multimarginal partial optimal transport (POT) problem between m discrete (unbalanced) measures with at most n supports. We first prove that we can obtain two equivalent forms of the multimarginal POT problem in terms of the multimarginal optimal transport problem via novel extensions of cost tensors. The first equivalent form is derived under the assumptions that the total masses of each measure are sufficiently close while the second equivalent form does not require any conditions on these masses but at the price of more sophisticated extended cost tensor. Our proof techniques for obtaining these equivalent forms rely on novel procedures of moving masses in graph theory to push transportation plan into appropriate regions. Finally, based on the equivalent forms, we develop an optimization algorithm, named the ApproxMPOT algorithm, that builds upon the Sinkhorn algorithm for solving the entropic regularized multimarginal optimal transport. We demonstrate that the ApproxMPOT algorithm can approximate the optimal value of multimarginal POT problem with a computational complexity upper bound of the order $\tilde{O}(m^3(n+1)^m/\varepsilon^2)$ where $\varepsilon > 0$ stands for the desired tolerance.

1 INTRODUCTION

The recent advances in computation of optimal transport (OT) (Villani, 2003; Cuturi, 2013; Lin et al., 2019b; Dvurechensky et al., 2018; Altschuler et al., 2017) have

brought new applications of optimal transport in machine learning and data science to the fore. Examples of these applications include generative models (Arjovsky et al., 2017; Tolstikhin et al., 2018; Gulrajani et al., 2017; Courty et al., 2017), unsupervised learning (Ho et al., 2017; Huynh et al., 2021), computer vision (Solomon et al., 2015; Nguyen et al., 2021), and other applications (Rolet et al., 2016; Peyré et al., 2016; Carrière et al., 2017). However, due to the marginal constraints of transportation plans, optimal transport is only defined between balanced measures, namely, measures with equal masses.

When measures are unbalanced, i.e., they can have different masses, there are two popular approaches for defining divergences between these measures. The first approach is unbalanced optimal transport (Chizat et al., 2015, 2016). The main idea of unbalanced optimal transport is to regularize the objective function of optimal transport based on certain divergences between marginal constraints of transportation plan and the masses of the input measures. Despite its favorable computational complexity (Pham et al., 2020) and practical applications (Schiebinger et al., 2019; Frogner et al., 2015; Janati et al., 2019; Balaji et al., 2020), the optimal transportation plan from unbalanced optimal transport is often non-trivial to interpret in practice.

The second approach for defining divergence between unbalanced measures is partial optimal transport (POT) (Caffarelli and McCann, 2010; Figalli, 2010), which was originally used to analyze partial differential equations. The idea of partial optimal transport is to constrain the total masses that will be transported between measures. It requires that the marginals of transport plan to be dominated by the corresponding measure. Due to these explicit constraints, the optimal transportation plan obtained from the partial optimal transport is more convenient to interpret than that achieved from the unbalanced optimal transport. As a consequence, partial optimal transport has begun to be employed in several machine learning applications recently. Examples of these applications include

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s). *Equal contribution.

computer graphics (Bonnel and Coeurjolly, 2019), graph neural networks (Sarlin et al., 2020), positive-unlabeled learning (Chapel et al., 2020), and partial covering (Kawano et al., 2021). To improve the scalability of partial optimal transport in these applications, the tree-sliced version of partial optimal transport has also been recently developed in (Le and Nguyen, 2021).

In this paper, we put the focus on POT, in particular its natural extension to deal with multiple (more than two) measures, namely multimarginal partial optimal transport (MPOT) (Kitagawa and Pass, 2015). Since MPOT has already been studied in previous works (e.g., (Kitagawa and Pass, 2015)), our main goal here is to provide an efficient tool for solving it. Specifically, we are interested in the computational aspect of MPOT in the discrete setting. The current literature offers two possible tools to solve MPOT: linear programming algorithms and the iterative Bregman projection (IBP) (Benamou et al., 2015). The former comes from the fact that MPOT problem is a linear program (LP) (see (2)), thus it can be solved by any linear programming algorithm. However, the number of variables and constraints in this case grow exponentially in the number of measures (i.e., n^m variables and $n^m + nm + 1$ constraints for m measures, each having n supports). Thus, the interior point methods for LPs may suffer from worst-case complexity bound $\mathcal{O}(n^{3m})$, while efficient network simplex algorithms used in the two-marginal case is inapplicable (see (Lin et al., 2020, Theorem 3.3) for the argument why the multimarginal formulation is not a minimum-cost flow problem when $m \geq 3$). On the other hand, the IBP algorithm is difficult to analyze. Inspired by current development in the computational tools (Lin et al., 2020) for multimarginal optimal transport (MOT) (Gangbo and Swiech, 1998; Pass, 2015), we show that these advances can be used to provide efficient tools to solve MPOT, via novel equivalent forms between MOT and MPOT.

Contribution. In this paper, we provide two equivalent forms for MPOT in terms of MOT via novel extensions of cost tensors. The first equivalent form is a natural development of the result in bi-marginal case established in (Chapel et al., 2020), but it requires the masses of all measures to be close to each other, which can be quite restrictive in general settings. To account for this limitation, we introduce the second equivalent form, which is more involved but free of additional conditions on the masses. The novel proof techniques for both equivalent forms come from the design of the extended cost tensors and the process of moving masses such that the marginal constraints of the transportation plan are still satisfied. Most of the challenges in our proofs are in the second equivalent form in which we have to design a more sophisticated

mass-moving procedure in order to push the masses into $m+1$ distant regions rather than neighbour regions in the first form. Finally, via these equivalent forms, we develop ApproxMPOT algorithm (which is built upon the Sinkhorn algorithm for solving the entropic regularized MOT (Lin et al., 2019a)) for approximating the solution of the MPOT problem. We prove that this algorithm has a complexity upper bound of the order $\tilde{\mathcal{O}}(m^3(n+1)^m/\varepsilon^2)$ for approximating the optimal MPOT cost where $\varepsilon > 0$ is the tolerance.

Organization. The paper is organized as follows. In Section 2, we provide background on optimal transport and multimarginal partial optimal transport. We then present two equivalent forms of multimarginal partial optimal transport in Section 3. Based on these equivalent forms, we develop ApproxMPOT algorithm for approximating the multimarginal POT and derive a complexity upper bound of that algorithm. We then present the experiments with the ApproxMPOT algorithm in Section 4 and conclude the paper in Section 5. Finally, we defer the missing proofs to the Appendices.

Notation. We let $[n]$ stand for the set $\{1, 2, \dots, n\}$ while \mathbb{R}_+^n stands for the set of all vectors in \mathbb{R}^n with non-negative entries. For a vector $\mathbf{x} \in \mathbb{R}^n$ and $p \in [1, \infty)$, we denote $\|\mathbf{x}\|_p$ as its ℓ_p -norm and $\text{diag}(\mathbf{x})$ as the diagonal matrix with \mathbf{x} on the diagonal. The natural logarithm of a vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}_+^n$ is denoted $\log \mathbf{a} = (\log a_1, \dots, \log a_n)$. $\mathbf{1}_n$ stands for a vector of length n with all of its components equal to 1. Let Σ_N be the histogram of N bins $\{\mathbf{p} \in \mathbb{R}_+^N : \sum_{i=1}^N p_i = 1\}$ and δ is the Dirac function. Next, given the dimension n and accuracy ε , the notation $a = \mathcal{O}(b(n, \varepsilon))$ means that $a \leq C \cdot b(n, \varepsilon)$ where C is independent of n and ε . Similarly, the notation $a = \tilde{\mathcal{O}}(b(n, \varepsilon))$ indicates the previous inequality may scale by a logarithmic function of n and ε . The entropy of a matrix \mathbf{X} is given by $H(\mathbf{X}) = \sum_{i,j=1}^n -X_{ij}(\log X_{ij} - 1)$. In general, we use \mathbf{X} for transportation matrix/tensor and $X_{(i,j)}$ or X_u for u in $[n+1]^m$ in the multimarginal case be its corresponding entries. We also define $\mathbf{X}_S = \{X_u : u \in S\}$, where S is a subset of $[n+1]^m$. Lastly, the cost matrix/tensor is denoted by \mathbf{C} , where its entries are denoted by C_v for $v \in [n]^m$.

2 PRELIMINARY

In this section, we first provide background for optimal transport between two discrete probability measures in Section 2.1. Then, we present (multimarginal) partial optimal transport between discrete (unbalanced) measures in Sections 2.2 and 2.3.

2.1 Optimal Transport

Assume that P and Q are discrete measures with at most $n \geq 1$ supports such that $P = \sum_{i=1}^n a_i \delta_{x_i}$ and $Q = \sum_{j=1}^n b_j \delta_{y_j}$. When P and Q have the same mass, i.e., $\|\mathbf{a}\|_1 = \|\mathbf{b}\|_1$, the optimal transport (OT) distance between P and Q admits the following form:

$$\text{OT}(P, Q) := \min_{\mathbf{X} \in \Pi(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{X} \rangle,$$

where \mathbf{C} is a cost matrix whose entries are distances between the supports of these distributions, $\mathbf{a} := (a_1, \dots, a_n)$, $\mathbf{b} := (b_1, \dots, b_n)$, and the set of admissible couplings $\Pi(\mathbf{a}, \mathbf{b})$ is given by:

$$\Pi(\mathbf{a}, \mathbf{b}) := \{\mathbf{X} \in \mathbb{R}_+^{n \times n} : \mathbf{X}\mathbf{1}_n = \mathbf{a}, \mathbf{X}^\top \mathbf{1}_n = \mathbf{b}\}. \quad (1)$$

2.2 Partial Optimal Transport

As indicated in equation (1), the OT formulation requires that P and Q have the same mass, i.e., P and Q are balanced measures. To account for the settings when P and Q are unbalanced measures, namely, $\|\mathbf{a}\|_1 \neq \|\mathbf{b}\|_1$, we consider in this section an approach named Partial Optimal Transport (POT), which focuses on optimally transporting only a fraction $0 \leq s \leq \min\{\|\mathbf{a}\|_1, \|\mathbf{b}\|_1\}$ of the mass. The set of all admissible couplings in that case is $\Pi^s(\mathbf{a}, \mathbf{b}) := \{\mathbf{X} \in \mathbb{R}_+^{n \times n} : \mathbf{X}\mathbf{1}_n \leq \mathbf{a}, \mathbf{X}^\top \mathbf{1}_n \leq \mathbf{b}, \mathbf{1}_n^\top \mathbf{X}\mathbf{1}_n = s\}$, and the formulation of the POT distance can be written as

$$\text{POT}(P, Q) := \min_{\mathbf{X} \in \Pi^s(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{X} \rangle.$$

At the first sight, the POT problem seems to be more demanding with the appearance of two inequality constraints in $\Pi^s(\mathbf{a}, \mathbf{b})$. However, we can tackle these inequality constraints by adding dummy points x_{n+1}, y_{n+1} and extending the cost matrix as follows (Chapel et al., 2020):

$$\bar{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{0}_n \\ \mathbf{0}_n^\top & A \end{pmatrix},$$

in which $A > 0$. We denote $\bar{\mathbf{a}} = [\mathbf{a}, \|\mathbf{b}\|_1 - s]$ and $\bar{\mathbf{b}} = [\mathbf{b}, \|\mathbf{a}\|_1 - s]$. The following result from (Chapel et al., 2020) demonstrates that solving the POT problem between P and Q is equivalent to solving an OT problem between two measures with mass $\|\mathbf{a}\|_1 + \|\mathbf{b}\|_1 - s$.

Proposition 1 ((Chapel et al., 2020)). *Under the assumption that $A > 0$, we have*

$$\min_{\mathbf{X} \in \Pi^s(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{X} \rangle = \min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{a}}, \bar{\mathbf{b}})} \langle \bar{\mathbf{C}}, \bar{\mathbf{X}} \rangle$$

and the optimal solution \mathbf{X}^* of the POT problem is exactly that of the OT problem, denoted by $\bar{\mathbf{X}}^*$, deprived from its last row and column.

The equivalent form in Proposition 1 indicates that by using the Sinkhorn algorithm for solving the entropic version of OT problem with cost matrix $\bar{\mathbf{C}}$ and the masses $\bar{\mathbf{a}}, \bar{\mathbf{b}}$, the computational complexity of approximating the POT between two measures P and Q will be at the order of $\mathcal{O}\left(\frac{(n+1)^2}{\varepsilon^2}\right)$ where $\varepsilon > 0$ stands for the desired tolerance.

2.3 Multimarginal Partial Optimal Transport

We now consider an extension of the partial optimal transport to multimarginal partial optimal transport (MPOT) when we have more than two measures. Assume that we are given m measures P_1, \dots, P_m with at most n supports that have weights $\mathbf{r}_1, \dots, \mathbf{r}_m$, respectively. The transport cost tensor/matrix denoted by $\mathbf{C} \in \mathbb{R}_+^{n \times \dots \times n}$ is also known in advance. Given a total mass of s that is needed to be transported, the multimarginal partial optimal transport wants to find a transport plan \mathbf{X}^* such that the i -th marginals of \mathbf{X}^* is a sub-measure of \mathbf{r}_i for all $i \in [m]$ and the total transport cost is minimized. In formula, \mathbf{X}^* is an optimal solution of the following problem:

$$\text{MPOT}(P_1, \dots, P_m) := \min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle, \quad (2)$$

where $\Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)$ is defined as:

$$\Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m) := \left\{ \mathbf{X} \in \mathbb{R}_+^{n \times \dots \times n} : c_k(\mathbf{X}) \leq \mathbf{r}_k, \forall k \in [m]; \sum_{v \in [n]^m} X_v = s \right\}, \quad (3)$$

with $[c_k(\mathbf{X})]_j = \sum_{i_\ell \in [n], \ell \neq k} X_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_m}$.

3 EQUIVALENT FORMS OF MULTIMARGINAL PARTIAL OPTIMAL TRANSPORT

In this section, we first present the challenges of extending the POT-OT equivalence in Proposition 1 to the multimarginal case. Then, we provide two approaches to link the multimarginal POT problem with its multimarginal OT (MOT) counterpart by cost tensor expansions. Subsequently, we will look into an algorithm to approximate the solution of multimarginal POT problem, which is inspired by two equivalent forms.

Our proofs for the equivalent forms of multimarginal POT mostly play with ‘‘layers’’ of the (cost or plan) tensors, which is defined as follow: for a subset S of $\{1, 2, \dots, m\}$, we denote by T_S the subset of tensor indices

$$T_S := \{(i_1, \dots, i_m) \in [n+1]^m : i_\ell = n+1 \Leftrightarrow \ell \in S\}.$$

For example, when $n = 2$ and $m = 3$ and $S = \{1, 2\}$, then $T_S = \{(3, 3, 1), (3, 3, 2)\}$. For all subsets S of the same cardinality k , the union of T_S (i.e., $\cup_{|S|=k} T_S$) forms a layer of order k (namely the k -th layer), and if $|S| = k$ then T_S is considered as a sublayer of this union. Thus, there are $m + 1$ layers in the m -dimensional hypercube.

3.1 The Three-Marginal Partial Optimal Transport

To begin with, we will investigate the simplest case of multimarginal POT, i.e., when $m = 3$. We will examine the structure of this problem under this simple case, and show the reason why we have two approaches for deriving equivalent forms for the general case.

In Figure 2, the red square (or cube) denotes the cost matrix (or tensor) \mathbf{C} , corresponding to the case $m = 2$ (or $m = 3$). When $m = 2$, by adding dummy points, the red square is extended by adding two green segments and a blue point corresponding to $C_{(n+1, n+1)}$. (Chapel et al., 2020) then show that the supports (i.e., positive entries) of the optimal transport plan lie on the red and green regions, with the total mass of the red region being equal to s , by solving two linear equations of the marginals. Given the uniform cost on the green parts, the OT problem reduces to minimizing its objective function in the red square. Thus, solving the bi-marginal POT is equivalent to solve a OT problem. For $m = 3$, there are two ways to extend the cost tensor to form two different systems of equations in order to prove that the total mass of the red part is equal to s .

The first approach is to build a cost tensor such that the supports of the transport plan lie only on the red and green regions in Figure 2, and the second approach is to build a cost tensor such that the supports lie only on the red and blue regions. In the language of “layers”, in the case $m = 3$, the green part and the blue part can be explicitly defined as $\{v \in T_S : S \subset \{1, 2, 3\}, |S| = 1\}$ and $\{v \in T_S : S \subset \{1, 2, 3\}, |S| = 2\}$, respectively.

The key idea of the equivalent forms is how to design mass-moving procedures inside the hyper-cube that keep the marginal constraints and decrease the objective function. After applying the procedure until we could not carry on further, the mass concentrates on the certain parts of the transport plan so that we can derive a system of linear equations to show that the total mass in the red region equals to s . Together with either the mass or the cost of the same layer of order $1, 2, \dots, m$ must be zero, it deduces the equivalence between the multimarginal POT and the multimarginal OT problems. Because the moving-mass procedure transports mass between neighbouring layers/regions, the closer between non-zero parts of the transport plan

are, the easier to design plan to push mass into those parts is. From this point, we could see that the first extension is the more natural extension of the bi-marginal case, when the non-zero parts (red and green) of the optimal transport plan are close to each other under the Hamming distance. That leads to a simple design of the cost tensor. The second extension requires novel technique to deal with, since the red (non-zero) parts of the optimal transport plan are far from the blue parts. That requires a more sophisticated mass moving procedure as well as the more complicated structure of the weights in the cost tensor.

3.2 The First Equivalent Form

Inspired by the extension technique in the bi-marginal case, we introduce the first equivalent form for the multimarginal POT problem. However, this equivalence comes at a cost of introducing additional conditions on the marginals that read

$$\Sigma_r := \sum_{i=1}^m \|\mathbf{r}_i\|_1 \geq (m-1)\|\mathbf{r}_k\|_1 + s, \quad \forall k \in [m]. \quad (4)$$

These conditions are necessary for the extended marginals to be non-negative (see Theorem 1). Note that these system of inequalities always hold for $m = 2$, and still hold for $m \geq 3$ in several cases, for example, the total masses of each measure are the same (i.e., $\|\mathbf{r}_i\|_1 = \|\mathbf{r}_j\|_1$ for all $i, j \in [m]$). We now state the first equivalent form under the conditions (4).

Theorem 1. *Assume that the conditions (4) are met. Let*

$$\bar{\mathbf{r}}_k^{(1)} = \left[\mathbf{r}_k, \frac{1}{m-1}\Sigma_r - \|\mathbf{r}_k\|_1 - \frac{1}{m-1}s \right]$$

be extended marginals and $\bar{\mathbf{C}}^{(1)} = [\bar{C}_v^{(1)} : v \in [n+1]^m]$, where

$$\bar{C}_v^{(1)} := \begin{cases} C_v & v \in T_\gamma \\ A_i & v \in T_S, |S| = i \quad \forall i \in [m] \end{cases}$$

be an extended cost tensor in which $0 = A_1 < A_2 < \dots < A_m$. Then, we find that

$$\min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle = \min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1^{(1)}, \dots, \bar{\mathbf{r}}_m^{(1)})} \langle \bar{\mathbf{C}}^{(1)}, \bar{\mathbf{X}} \rangle. \quad (5)$$

Moreover, assume that $\bar{\mathbf{X}}^*$ is a minimizer of the multimarginal OT problem, then $\mathbf{X}^* := \bar{\mathbf{X}}_{[n]}^*$ is an optimal solution of the multimarginal POT problem.

Example 1. *To illustrate the result of Theorem 1, we assume that $m = 3$, i.e., 3-marginal setting, and $\|\mathbf{r}_1\|_1 = \|\mathbf{r}_2\|_1 = \|\mathbf{r}_3\|_1 = 1$. Under this setting, the condition (4) is satisfied. For simplicity, we set $n = 2$, meaning each marginal has two support points. Based*

on Theorem 1, the extended cost tensor $\bar{\mathbf{C}}^{(1)}$ takes the following form: $\bar{C}_{u_1}^{(1)} = 0, \bar{C}_{u_2}^{(1)} = 1, \bar{C}_{u_3}^{(1)} = 2$, in which

$$\begin{aligned} u_1 \in U_1 &:= \left\{ (3, 1, 1), (3, 1, 2), (3, 2, 1), (3, 2, 2), \right. \\ &\quad (1, 3, 1), (1, 3, 2), (2, 3, 1), (2, 3, 2), \\ &\quad \left. (1, 1, 3), (1, 2, 3), (2, 1, 3), (2, 2, 3) \right\}; \\ u_2 \in U_2 &:= \left\{ (1, 3, 3), (2, 3, 3), (3, 1, 3), \right. \\ &\quad \left. (3, 2, 3), (3, 3, 1), (3, 3, 2) \right\}; \\ u_3 \in U_3 &:= \left\{ (3, 3, 3) \right\}. \end{aligned}$$

Remark 1. Theorem 1 requires a simple condition that the sequence (A_i) needs to be monotone increasing. In comparison with (Chapel et al., 2020), they use the sequence $(0, A)$ for the bi-marginal case. That resemblance partly explains why the first equivalence is the more natural extension of the bi-marginal case. The condition (4) could be met in some situations, i.e., the values of $\|\mathbf{r}_i\|_1$ are the same. However, in general, it is quite restrictive, since the condition requires that the magnitude of those positive measures are close to each other.

In order to prove the results of both Theorems 1 and 2, we need Lemma 1 which plays a key role in our proofs. More specifically, this lemma shows multiple ways to transport mass inside the multimarginal plan \mathbf{X} such that the marginal constraints of \mathbf{X} still hold. Since Lemma 1 works on the hypercube $\mathcal{C}_k = \{0, 1\}^k$, we start with some notations and definitions of the hypercube.

Definition 1 (Hyper-cube and its graph). Let $\mathcal{C}_k = \{0, 1\}^k = \{(i_1, \dots, i_k) : i_\ell \in \{0, 1\}\}$ be the hyper-cube of dimension k . The graph on hyper-cube \mathcal{C}_k is defined as follow: for $u = (u_1, \dots, u_k), v = (v_1, \dots, v_k) \in \mathcal{C}_k$, u and v are connected if and only if there exists only one index j such that $1 \leq j \leq k$ and $u_j \neq v_j$. In the language of distance, the Hamming distance between u and v is equal to 1.

Definition 2 (Hyper-rectangle and its graph). Let $m \geq k$ be two positive integers. Assume that $u = (u_1, \dots, u_m), v = (v_1, \dots, v_m) \in \mathbb{Z}^m$ and $|\{u_i \neq v_i, 1 \leq i \leq m\}| = k$. The hyper-rectangle which has u and v as two opposite vertices is defined as follow:

$$\mathcal{P}(u, v) = \{(w_1, \dots, w_m) : w_i \in \{u_i, v_i\} \forall i \in [m]\}.$$

Similarly, the graph on the hyper-rectangle is defined as follows : $w = (w_1, \dots, w_m)$ and $\tilde{w} = (\tilde{w}_1, \dots, \tilde{w}_m)$ are connected if and only if there exists only one index ℓ such that $w_\ell \neq \tilde{w}_\ell$.

In graph theory, the graph of hyper-rectangle is the same as the graph of the hyper-cube. Hence, the bellow

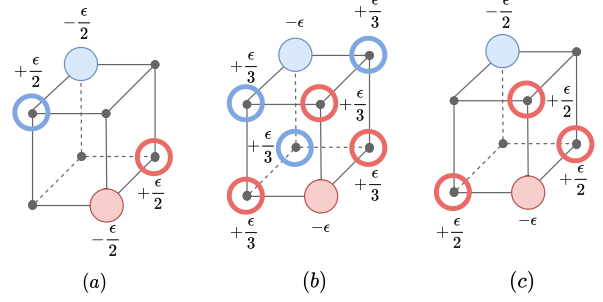


Figure 1: Visualization of mass-moving procedures in Lemma 1 for $k = 3$. Blue-filled circles represent vertices $\mathbf{0}$ and red-filled circles represent vertices $\mathbf{1}$, while the unfilled circles depict the neighbors of the color-corresponding filled nodes. The number, that is next to the node, i.e., ϵ , is the mass that node gains or loses in the mass-moving procedure.

lemma for the hyper-cube could be applied to the hyper-rectangle as well.

Lemma 1 (Preservative Mass Transportation on Hypercube). Let $\mathcal{C}_k = \{0, 1\}^k$ be a hyper-cube in the k -dimensional space. For each vertex v of the cube \mathcal{C}_k , we assign a mass number M_v . Assume that $0 < \epsilon < \min\{M_{\mathbf{0}_k}, M_{\mathbf{1}_k}\}$, then each of the following operations will keep the quantity

$$\sum_{v \in \text{one face of } \mathcal{C}_k} M_v$$

unchanged for any faces of the cube.

(a) We move $\frac{\epsilon}{2}$ mass from vertex $u = \mathbf{0}_k$ to its neighbours \tilde{u} and we move $\frac{\epsilon}{2}$ mass from vertex $v = \mathbf{1}_k$ to its neighbour \tilde{v} such that the edge (u, \tilde{u}) is parallel to the edge (v, \tilde{v}) .

(b) We move $\frac{\epsilon}{k}$ mass from vertices $\mathbf{0}_k$ and $\mathbf{1}_k$ to each of their neighbours (ϵ mass in total).

(c) We take away ϵ and $\frac{\epsilon}{k-1}$ mass from $\mathbf{1}_k$ and $\mathbf{0}_k$, respectively. We add $\frac{\epsilon}{k-1}$ to all k neighbours of $\mathbf{1}_k$.

An illustration of Lemma 1 is in Figure 1, and the proof is deferred to Appendix C.1.

Proof Sketch of Theorem 1. The full proof of Theorem 1 can be found in Appendix A. A key step in the proofs of equivalent forms is to show that $\sum_{v \in T_\gamma} \bar{X}_v^* = s$, i.e., the total mass at the portion of $\bar{\mathbf{X}}^*$ corresponding to the original cost tensor (the zeroth layer) is the total mass we want to transport in the original partial problem. To see this, we start by showing that there is some mass in the zeroth layer, i.e., there exists $u \in T_\gamma$ such that $\bar{X}_u^* > 0$. Next, we prove by contradiction that the mass of $\bar{\mathbf{X}}^*$ only lies at

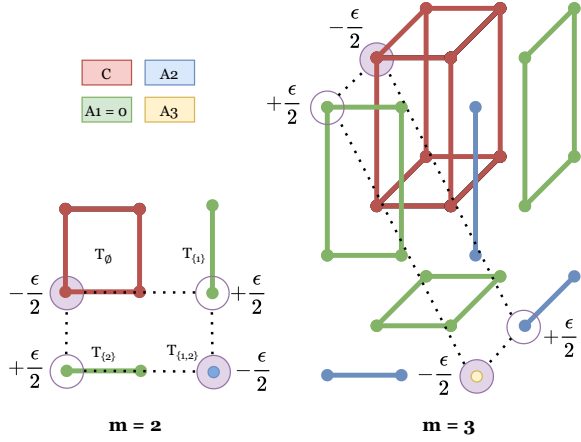


Figure 2: An illustration for mass-moving procedures in the proof of the first equivalent form in two cases $m = 2$ and $m = 3$, where the red part corresponds to the original supports of the cost plan and others are the extended, $+\frac{\epsilon}{2}$ means adding $\frac{\epsilon}{2}$ mass to that node, and $-\frac{\epsilon}{2}$ means take away $\frac{\epsilon}{2}$ mass from that node.

the zeroth layer and the first layer (which correspond to the red region and the green region in Figure 2, respectively), i.e., $\bar{X}_v^* = 0$ if $v \notin T_\gamma \cup (\cup_{j=1}^m T_{\{j\}})$. If it is not the case, then we can construct a preservative mass-moving procedure (see Lemma 1 and Figure 2) involving the zeroth layer (in particular, the location $u \in T_\gamma$ with positive mass), and the cost design will invalidate the optimality of \bar{X}^* . After that, the aforementioned key statement can be attained using some algebraic transformations on marginal constraints. Consequently, we get

$$\begin{aligned} \min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1, \dots, \bar{\mathbf{r}}_m)} \langle \bar{\mathbf{C}}^{(1)}, \bar{\mathbf{X}} \rangle &= \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^* \rangle \\ &\geq \min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle, \end{aligned}$$

and we prove that the equality by showing the inverse inequality, which is done via some construction of $\bar{\mathbf{X}}^*$ based on \mathbf{X}^* . \square

3.3 The Second Equivalent Form

The first equivalent form requires the total masses $\|\mathbf{r}_i\|_1$ where $1 \leq i \leq m$ to satisfy the condition (4), which can be undesirable under general settings of unbalanced measures. To circumvent this limitation of the first equivalence, we propose another equivalent form to the multimarginal POT problem, where the extended cost tensor is more sophisticated and less sparse than that of the first equivalent form. However, no additional conditions are needed for the second equivalence, except the conditions $s \leq \|\mathbf{r}_j\|_1$ for all $j \in [m]$, which are apparently essential in the original formulation of multimarginal POT.

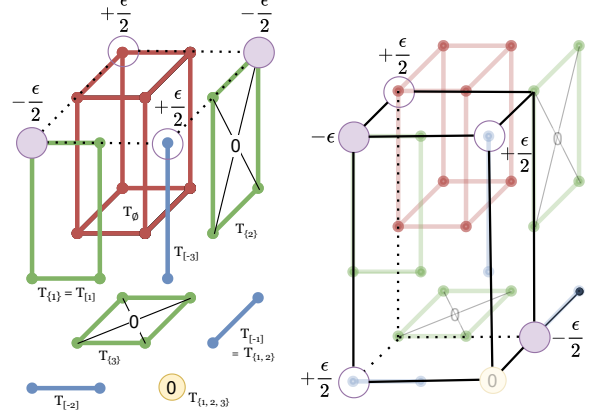


Figure 3: An illustration for mass-moving procedures in the proof of the second equivalent form when $m = 3$.

Theorem 2. Let $\{D_i\}_{i=0}^m$ be a sequence satisfying $D_0 = \|\mathbf{C}\|_\infty, D_{m-1} = 0, D_m > 0$. Furthermore, suppose that (D_0, D_1, D_2) is a concave sequence when $m = 3$, while for $m \geq 4$, we assume

$$\Delta_i^{(2)} \leq (m-1-i)\Delta_{i+1}^{(2)} \leq 0, \quad \forall i \in [m-3],$$

where $\Delta_i^{(2)} := D_{i+1} + D_{i-1} - 2D_i$. Let

$$\bar{\mathbf{r}}_k^{(2)} = \left[\mathbf{r}_k, \sum_{i \neq k} \|\mathbf{r}_i\|_1 - (m-1)s \right]$$

be extended marginals and $\bar{\mathbf{C}}^{(2)} = [\bar{C}_v^{(2)} : v \in [n+1]^m]$, where

$$\bar{C}_v^{(2)} = \begin{cases} C_v & v \in T_\gamma \\ D_i & v \in T_S, |S| = i \quad \forall i \in [m] \end{cases}$$

be an extended cost tensor. Then,

$$\min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle = \min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1^{(2)}, \dots, \bar{\mathbf{r}}_m^{(2)})} \langle \bar{\mathbf{C}}^{(2)}, \bar{\mathbf{X}} \rangle. \quad (6)$$

Similar to Theorem 1, assume that $\bar{\mathbf{X}}^*$ is a minimizer of the multimarginal OT problem, then $\mathbf{X}^* := \bar{\mathbf{X}}_{[n]}^*$ is an optimal solution of the multimarginal POT problem.

Proof Sketch of Theorem 2. For the full proof, the readers can refer to Appendix B. Note that each layer can be decomposed into corresponding sublayers (e.g., the first layer $|S| = 1$ consists of three sublayers $T_{\{1\}}, T_{\{2\}}, T_{\{3\}}$ corresponding to the green region in Figure 3). The first step is to show that the supports of the optimal transportation plan (called *optimal supports*) must lie in a sequence of nested sublayers when their orders are in $[m-2]$; if not, use a mass-moving procedure (depicted on Figure 3, left) built on nodes in two sublayers of the same layer to lower the objective function. This is due to the concavity of the

constructed cost sequence $\{D_i\}_{i=1}^{m-2}$ (see Theorem 2). The next step is to show that the optimal supports must not lie in the layers of order $1, 2, \dots, m-2$, which relies on the procedure between one of the layers of order $1, \dots, m-2$ and the layer of order $m-1$ (see Figure 3, right), and also due to the noted concavity and $D_{m-1} = 0$. Thus, the optimal supports must lie in the original layer (i.e., the red T_γ) and the two last (blue and yellow) layers. Finally, simple derivations show that the total optimal mass on the original layer is s , and subsequent reasoning is similar to the last step in the proof of Theorem 1. \square

Remark 2. *Theorem 2 requires some conditions on the second order of difference of the sequence (D_i) that trade off for having no conditions on the \mathbf{r}_i . Given the sequence $(\Delta_i^{(2)})$ and let $\Delta_i^{(1)} := D_{i+1} - D_i$, we could build up the sequence (D_i) as follows:*

$$\begin{aligned} D_i &= \Delta_{i-1}^{(1)} + \dots + \Delta_0^{(1)} + D_0 \\ &= \Delta_{i-1}^{(2)} + 2\Delta_{i-2}^{(2)} + \dots + (i-1)\Delta_1^{(2)} + i\Delta_0^{(1)} + D_0. \end{aligned}$$

When all the chosen $\Delta_j^{(2)}$ satisfying the conditions $\Delta_j^{(2)} \leq (m-1-j)\Delta_{j+1}^{(2)} \leq 0$ (e.g., $\Delta_j^{(2)} = -(m-1-i)!$), we choose $\Delta_0^{(1)} = -\frac{\sum_{i=1}^{m-2} (m-1-i)\Delta_i^{(2)} + D_0}{m-1}$, which leads to the fact that $D_{m-1} = 0$. Then, the sequence (D_i) satisfies all the required conditions. Since the sequence (D_i) is concave, its minimum will be at the end terms of the sequence, which are D_{m-1} and D_0 . This means that other D_i cannot be negative.

Example 2. *Similar to Example 1, we illustrate the construction of the extended cost tensor in Theorem 2 under the 3-marginal setting, i.e., $m = 3$, and $\|r_1\|_1 = \|r_2\|_1 = \|r_3\|_1 = 1$. For simplicity, we set $n = 2$, meaning that each marginal has two support points. Based on Theorem 2, the extended cost tensor $\bar{\mathbf{C}}^{(2)}$ takes the following form*

$$\bar{C}_u^{(2)} = \begin{cases} D_1, & u \in U_1 \\ 0, & u \in U_2, \\ D_3, & u \in U_3 \end{cases}$$

where we choose $D_1 > D_0 = \max_{u \in [n]^m} C_u$ and $D_3 > 0$ while U_1, U_2, U_3 are defined as in Example 1. Here, the sequence D_0, D_1, D_2 is already concave.

3.4 Algorithmic Developments

In this section, we briefly derive algorithmic procedure to approximate the multimarginal partial optimal transport problem based on the previous equivalences, as well as the computational complexity of the approximating algorithms. Algorithm 1 is the approximating algorithm based on the Sinkhorn procedure (namely

Algorithm 1 ApproxMPOT

Parameters: $\mathbf{C}, \{\mathbf{r}_i\}_{i=1}^m, s, \eta, \varepsilon$

Extend the cost tensor \mathbf{C} into $\bar{\mathbf{C}}$ and $\{\mathbf{r}_i\}_{i=1}^m$ into $\{\bar{\mathbf{r}}_i\}_{i=1}^m$ according to Theorem 1 or Theorem 2

Compute $\bar{\mathbf{X}}^k = \text{SinkhornMOT}(\bar{\mathbf{C}}, \eta, \{\bar{\mathbf{r}}_i\}_{i=1}^m, \varepsilon)$, where k is the number of Sinkhorn iterations

Let $\mathbf{X}^k = \bar{\mathbf{X}}^k[1:n, \dots, 1:n]$

return \mathbf{X}^k

SinkhornMOT) to solve entropic-regularized multimarginal OT given in (Lin et al., 2019a, Algorithm 3). Interestingly, the computational complexity of the multimarginal POT can also be derived from that of multimarginal OT. Before stating that complexity result, we first define the notion of ε -approximated multimarginal partial transportation plan.

Definition 3 (ε -approximation). *The tensor $\hat{\mathbf{X}} \in \mathbb{R}_+^{n^m}$ is called an ε -approximated multimarginal partial transportation plan if $c_k(\hat{\mathbf{X}}) \leq r_k$ for any $k \in [m]$ and the following inequality holds true,*

$$\langle \mathbf{C}, \hat{\mathbf{X}} \rangle \leq \langle \mathbf{C}, \mathbf{X}^* \rangle + \varepsilon,$$

where \mathbf{X}^* is defined as an optimal solution of the multimarginal POT problem (2).

Given Definition 3, we have the following proposition about the computational complexity of ApproxMPOT algorithm for approximating the multimarginal POT problem.

Proposition 2. *Algorithm 1 returns an ε -approximated multimarginal partial transportation plan \mathbf{X}^k within $\mathcal{O}\left(\frac{m^3(n+1)^m \|\bar{\mathbf{C}}\|_\infty^2 \log(n+1)}{\varepsilon^2}\right)$ arithmetic operations where $\bar{\mathbf{C}}$ is a given cost tensor in either Theorem 1 or Theorem 2.*

The proof of Proposition 2 is in Appendix C.2.

4 EMPIRICAL STUDY

A Simple Illustration of Robustness. In this experiment, we empirically verify that MPOT is more robust to outliers compared to MOT, in the sense the optimal MPOT cost is less sensitive in the appearance of noisy marginal support. We consider three empirical measures of 10 supports sampled from $\mathcal{N}((0, 0), I_2)$, $\mathcal{N}((1, 1), I_2)$, and $\mathcal{N}((-1, 1), I_2)$ respectively, and gradually inject $n_0 \in \{1, 2, 3, 4, 5\}$ noisy support points drawn from faraway Gaussians $\mathcal{N}((0, 5), I_2)$, $\mathcal{N}((5, 5), I_2)$, and $\mathcal{N}((-5, 5), I_2)$, respectively. The weights on the total $10 + n_0$ support points are set uniformly. The corresponding MOT and MPOT costs (with squared Euclidean ground metric) are reported in Figure 4, which highlights the expected behavior.

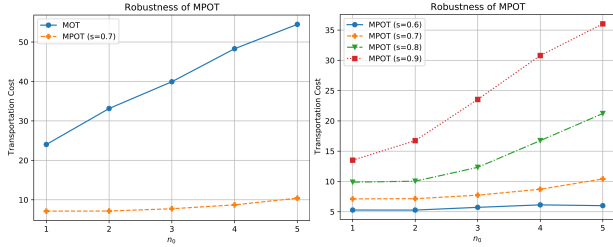


Figure 4: Robustness of the transportation cost of MPOT compared to MOT when injecting noise into three marginals. *Left*: MPOT and MOT in comparison. *Right*: MPOT with different prespecified mass $s \in \{0.6, 0.7, 0.8, 0.9\}$.

Partial Barycenter. It is known that the multimarginal partial optimal transport is equivalent to the partial barycenter problem (Kitagawa and Pass, 2015, Proposition 1.1). In this experiment, we investigate the robust behavior of MPOT by using it to compute the (partial) barycenter of three corrupted Gaussian measures. Specifically, we present each Gaussian measure by a histogram over 100 support points, where the masses on these points come from a mixture of two Gaussian distribution with weights 0.9 and 0.1 (the former corresponds to the true underlying distribution, while the latter is the noise - this simulates Huber’s 0.1-contamination model). The detail can be found in the Figure 5, in which we plot three barycenters: one corresponding to the standard optimal transport, one to the partial optimal transport, and one coming from solving a MPOT problem (Benamou et al., 2015, Section 5.3). These barycenters are computed using convex solvers (Agrawal et al., 2018) on the corresponding entropic-regularized formulations with $\eta = 1$ for smoother visualization. It is apparent that the histograms from the partial barycenter problem and from MPOT resemble each other, and are not affected by outliers (while the OT-based histogram is).

Empirical Convergence of ApproxMPOT. Next, we take a look into the convergence of our approximating algorithm for different regularization values. Specifically, we set $m = 3, n = 10, s = 0.8$ and run the algorithm for $\eta \in \{0.01, 0.1, 1\}$ then plot the objective values over iterations. The optimal value of unregularized problem is computed by a convex solver (Agrawal et al., 2018). As we can see from the plot in Figure 6, small η converges slower, but large η may result in an inaccurate approximation.

5 CONCLUSION

In this paper, we study the multimarginal partial optimal transport problem between $m \geq 2$ (unbalanced)

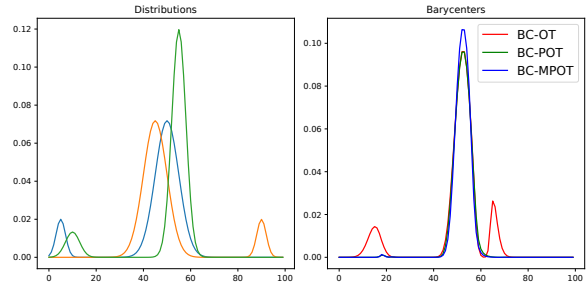


Figure 5: Robustness of multimarginal partial optimal transport. *Left*: three corrupted Gaussian histograms of $0.9 \cdot \mathcal{N}(50, 25) + 0.1 \cdot \mathcal{N}(5, 4)$ (blue), $0.9 \cdot \mathcal{N}(45) + 0.1 \cdot \mathcal{N}(90, 4)$ (orange), $0.9 \cdot \mathcal{N}(55, 9) + 0.1 \cdot \mathcal{N}(10, 9)$ (green). *Right*: the standard OT barycenter (red), the partial OT barycenter (green), and the barycenter computed by MPOT (blue). The partial algorithms are run with total masses of 0.8.

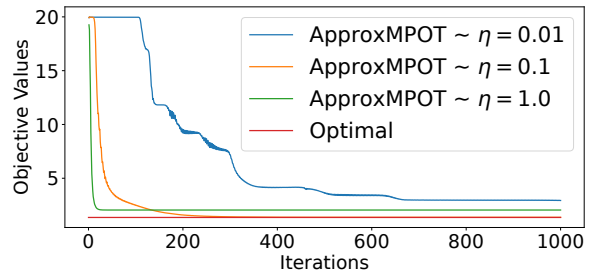


Figure 6: Convergence of ApproxMPOT. The red line is the optimal value of the unregularized problem, while three other lines correspond to different runs of the algorithm for solving the regularized problems with different values of η .

discrete measures with at most n supports. We first establish two equivalent forms of the multimarginal POT problem in terms of the multimarginal OT problem via novel extensions of the cost tensors. Our proof techniques are based on sophisticated procedures of designing extended cost tensors and of moving masses such that the transportation plan will put its mass into appropriate regions. Based on these equivalent forms, we then develop an optimization algorithm, named ApproxMPOT algorithm, for approximating the multimarginal POT with a computational complexity upper bound of the order $\tilde{O}(m^3(n+1)^m/\varepsilon^2)$ where $\varepsilon > 0$ is the tolerance.

References

- A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018. (Cited on page 8.)
- J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in neural information processing systems*, pages 1964–1974, 2017. (Cited on page 1.)
- M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223, 2017. (Cited on page 1.)
- Y. Balaji, R. Chellappa, and S. Feizi. Robust optimal transport with applications in generative modeling and domain adaptation. In *NeurIPS*, 2020. (Cited on page 1.)
- J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015. (Cited on pages 2 and 8.)
- N. Bonneel and D. Coeurjolly. SPOT: Sliced partial optimal transport. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019. (Cited on page 2.)
- L. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Annals of Mathematics*, pages 673–730, 2010. (Cited on page 1.)
- M. Carrière, M. Cuturi, and S. Oudot. Sliced Wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 1–10, 2017. (Cited on page 1.)
- L. Chapel, M. Z. Alaya, and G. Gasso. Partial optimal transport with applications on positive-unlabeled learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 2903–2913. Curran Associates, Inc., 2020. (Cited on pages 2, 3, 4, and 5.)
- L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulation. *Arxiv preprint Arxiv: 1508.05216*, 2015. (Cited on page 1.)
- L. Chizat, G. Peyré, B. Schmitzer, and F. Vialard. Scaling algorithms for unbalanced transport problems. *ArXiv Preprint: 1607.05816*, 2016. (Cited on page 1.)
- N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. (Cited on page 1.)
- M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013. (Cited on page 1.)
- P. Dvurechensky, A. Gasnikov, and A. Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn’s algorithm. In *International conference on machine learning*, pages 1367–1376, 2018. (Cited on page 1.)
- A. Figalli. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, page 533–560, 2010. (Cited on page 1.)
- C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, pages 2053–2061, 2015. (Cited on page 1.)
- W. Gangbo and A. Swiech. Optimal maps for the multidimensional Monge-Kantorovich problem. *Communications on Pure and Applied Mathematics*, 51(1):23–45, 1998. (Cited on page 2.)
- I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017. (Cited on page 1.)
- N. Ho, X. Nguyen, M. Yurochkin, H. Bui, V. Huynh, and D. Phung. Multilevel clustering via Wasserstein means. In *ICML*, 2017. (Cited on page 1.)
- V. Huynh, N. Ho, N. Dam, L. Nguyen, M. Yurochkin, H. Bui, and D. Phung. On efficient multilevel clustering via Wasserstein distances. *Journal of Machine Learning Research*, pages 1–43, 2021. (Cited on page 1.)
- H. Janati, M. Cuturi, and A. Gramfort. Wasserstein regularization for sparse multi-task regression. In *AISTATS*, 2019. (Cited on page 1.)
- K. Kawano, S. Koide, and K. Otaki. Partial Wasserstein covering. *Arxiv preprint Arxiv: 2106.00886*, 2021. (Cited on page 2.)
- J. Kitagawa and B. Pass. The multi-marginal optimal partial transport problem. *Forum of Mathematics, Sigma*, 2015. (Cited on pages 2 and 8.)
- T. Le and T. Nguyen. Entropy partial transport with tree metrics: Theory and practice. In *AISTATS*, 2021. (Cited on page 2.)
- T. Lin, N. Ho, M. Cuturi, and M. I. Jordan. On the complexity of approximating multimarginal optimal transport. *arXiv preprint arXiv:1910.00152*, 2019a. (Cited on pages 2, 7, and 17.)
- T. Lin, N. Ho, and M. I. Jordan. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. *ICML*, 2019b. (Cited on page 1.)

- T. Lin, C. Fan, N. Ho, M. Cuturi, and M. I. Jordan. Projection robust Wasserstein distance and Riemannian optimization. *arXiv preprint arXiv:2006.07458*, 2020. (Cited on page 2.)
- T. Nguyen, Q.-H. Pham, T. Le, T. Pham, N. Ho, and B.-S. Hua. Point-set distances for learning representations of 3D point clouds. In *ICCV*, 2021. (Cited on page 1.)
- B. Pass. Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015. (Cited on page 2.)
- G. Peyré, M. Cuturi, and J. Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *ICML*, pages 2664–2672, 2016. (Cited on page 1.)
- K. Pham, K. Le, N. Ho, T. Pham, and H. Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *ICML*, 2020. (Cited on page 1.)
- A. Rolet, M. Cuturi, and G. Peyré. Fast dictionary learning with a smoothed Wasserstein loss. In *AISTATS*, pages 630–638, 2016. (Cited on page 1.)
- P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. (Cited on page 2.)
- G. Schiebinger et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176:928–943, 2019. (Cited on page 1.)
- J. Solomon, F. Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. In *SIGGRAPH*, 2015. (Cited on page 1.)
- I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schölkopf. Wasserstein auto-encoders. In *International Conference on Learning Representations (ICLR 2018)*, 2018. (Cited on page 1.)
- C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003. (Cited on page 1.)

Supplementary Material: On Multimarginal Partial Optimal Transport: Equivalent Forms and Computational Complexity

In this supplementary material, we firstly provide the proof of Theorem 1 in Appendix A while the proof of Theorem 2 is subsequently presented in Appendix B. Finally, Appendix C is devoted to show proofs for Lemma 1 and Proposition 2.

A PROOF OF THEOREM 1

To prove $A = B$, we prove that $A \geq B$ and $B \geq A$, where A and B are the transportation cost of the multimarginal POT and multimarginal OT, respectively. Denote by $\bar{\mathbf{X}}^*$ the solution of the multimarginal OT problem $\min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1^{(1)}, \dots, \bar{\mathbf{r}}_m^{(1)})} \langle \bar{\mathbf{C}}^{(1)}, \bar{\mathbf{X}} \rangle$. We divide the proof of Theorem 1 into four key steps.

Step 1: We first prove that $\sum_{v \in T_\ell} \bar{X}_v^* \geq s$. Let $T_{(\ell)} := \{v \in [n+1]^m : v_\ell = n+1\}$. For each extended marginal $\bar{\mathbf{r}}_\ell^{(1)}$, considering its last element, we get

$$\sum_{v \in T_{(\ell)}} \bar{X}_v^* = \frac{1}{m-1} \Sigma_{\mathbf{r}} - \|\mathbf{r}_\ell\|_1 - \frac{1}{m-1} s.$$

Summing over all $\ell \in [m]$ leads to

$$\sum_{\ell=1}^m \sum_{v \in T_{(\ell)}} \bar{X}_v^* = \frac{m}{m-1} \Sigma_{\mathbf{r}} - \Sigma_{\mathbf{r}} - \frac{m}{m-1} s = \frac{1}{m-1} (\Sigma_{\mathbf{r}} - ms).$$

Since $\sum_{v \in \bigcup_{\ell=1}^m T_{(\ell)}} \bar{X}_v^* \leq \sum_{\ell=1}^m \sum_{v \in T_{(\ell)}} \bar{X}_v^*$, it follows that

$$\sum_{v \in \bigcup_{\ell=1}^m T_{(\ell)}} \bar{X}_v^* \leq \frac{1}{m-1} (\Sigma_{\mathbf{r}} - ms).$$

Moreover, we have $[n]^m = [n+1]^m \setminus (\bigcup_{j=1}^m T_{(j)})$. It means that

$$\begin{aligned} \sum_{v \in [n]^m} \bar{X}_v^* &= \sum_{v \in [n+1]^m} \bar{X}_v^* - \sum_{v \in \bigcup_{j=1}^m T_{(j)}} \bar{X}_v^* \\ &\geq \frac{1}{m-1} (\Sigma_{\mathbf{r}} - s) - \frac{1}{m-1} (\Sigma_{\mathbf{r}} - ms) \\ &= s. \end{aligned}$$

Hence, there exists $u \in [n]^m$ such that $\bar{X}_u^* > 0$.

Step 2: We prove that if $\bar{X}_v^* > 0$ then $v \in T_\ell \cup (\bigcup_{j=1}^m T_{(j)})$. Assume the contrary, which means that there exists $v = (v_1, \dots, v_m)$ in some T_S with $|S| = j > 1$ such that $\bar{X}_v^* > 0$. WLOG, we assume that $v_1 = v_2 = n+1$. We build an hyper-rectangle where

$$\begin{aligned} u &= (u_1, u_2, u_3, \dots, u_m) \in T_\ell, \\ v &= (n+1, n+1, v_3, \dots, v_m), \end{aligned}$$

correspond to $\mathbf{0}_k$ and $\mathbf{1}_k$, respectively. Let $\epsilon = \min \{\bar{X}_u^*, \bar{X}_v^*\}$ and apply the procedure in Lemma 1(a) with neighbors

$$\begin{aligned}\tilde{u} &= (n+1, u_2, u_3, \dots, u_m), \\ \tilde{v} &= (u_1, n+1, v_3, \dots, v_m).\end{aligned}$$

Then the total cost is changed by

$$-\frac{\epsilon}{2}(C_u^{(1)} - A_1 + A_j - A_{j-1}),$$

which is negative (since $C_u^{(1)} \geq 0, A_1 = 0, A_j \geq A_{j-1}$), contradictory to the optimality of $\bar{\mathbf{X}}^*$.

Step 3: We prove that $\sum_{v \in T_\gamma} \bar{X}_v^* = s$. For each extended marginal $\bar{\mathbf{r}}_\ell^{(1)}$, considering its first m -th elements, we get

$$\sum_{v \in [n+1]^m \setminus T_{\{\ell\}}} \bar{X}_v^* = \|\mathbf{r}_\ell\|_1.$$

Moreover, the result in Step 2 indicates that $\bar{X}_v^* = 0$ for all $v \notin T_\gamma \cup (\cup_{j=1}^m T_{\{j\}})$. Thus,

$$\sum_{v \in T_\gamma} \bar{X}_v^* + \sum_{v \in T_{\{j\}}, j \neq \ell} \bar{X}_v^* = \|\mathbf{r}_\ell\|_1.$$

Summing over all $\ell \in [m]$ leads to

$$m \sum_{v \in T_\gamma} \bar{X}_v^* + (m-1) \sum_{j=1}^m \sum_{v \in T_{\{j\}}} \bar{X}_v^* = \Sigma_{\mathbf{r}}.$$

Note that in Step 1 we have obtained

$$\sum_{\ell=1}^m \sum_{v \in T_{\{\ell\}}} \bar{X}_v^* = \frac{1}{m-1}(\Sigma_{\mathbf{r}} - ms).$$

From the above equations, we deduce that $\sum_{v \in T_\gamma} \bar{X}_v^* = s$.

Step 4: We prove the claimed statement. Let $\bar{\mathbf{X}}_{[n]^m}^*$ is the sub-tensor of $\bar{\mathbf{X}}^*$ corresponding to the first m components in each dimension. By construction, $c_k(\bar{\mathbf{X}}_{[n]^m}^*) \leq \mathbf{r}_k$ and from Step 3, $\sum_{v \in [n]^m} (\bar{\mathbf{X}}_{[n]^m}^*)_v = s$. Since for any $v \notin [n]^m$, either the corresponding cost $C_v^{(1)}$ is zero (note that $A_1 = 0$) or the corresponding optimal mass \bar{X}_v^* is zero (see Step 2), we obtain

$$\langle \bar{\mathbf{C}}^{(1)}, \bar{\mathbf{X}}^* \rangle = \langle \mathbf{C}, \bar{\mathbf{X}}_{[n]^m}^* \rangle.$$

This means that

$$\min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1, \dots, \bar{\mathbf{r}}_m)} \langle \bar{\mathbf{C}}^{(1)}, \bar{\mathbf{X}} \rangle = \langle \mathbf{C}, \bar{\mathbf{X}}_{[n]^m}^* \rangle \geq \min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle.$$

The next step is to prove the inverse inequality in order to deduce the equality. Let

$$\mathbf{X}^* = \arg \min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle,$$

we are going to expand \mathbf{X}^* to become the optimal plan $\tilde{\mathbf{X}}^*$ of the MOT problem, where $\tilde{X}_v^* = X_v^*$ for all $v \in T_\gamma$. We also know from Step 2 that $\tilde{X}_v^* = 0$ for all $v \notin \cup_{j=1}^m T_{\{j\}}$. The problem thus boils down to define \tilde{X}_v^* for all $v \in \cup_{j=1}^m T_{\{j\}}$ satisfying the marginal constraints.

Note that if we view each element of the set $\{\tilde{X}_v^* : v \in T_{\{k\}}\}$ as a tensor of $[n]^{m-1}$, then it has $m-1$ marginals which are denoted by S_k^ℓ for $\ell \in [m]$ and $\ell \neq k$. In formula, we have

$$\begin{aligned} S_k^\ell &= (S_{k,1}^\ell, \dots, S_{k,n}^\ell); \\ S_{k,i}^\ell &= \sum_{v: v \in T_{\{k\}}, v_\ell = i} \tilde{X}_v^*, \quad i \in [n]. \end{aligned}$$

The ℓ -th marginal constraint of $\tilde{\mathbf{X}}^*$ is written as

$$\begin{aligned} (\bar{\mathbf{r}}_\ell)_{n+1} &= \sum_{v \in T_{\{\ell\}}} \tilde{X}_v^* = \|S_\ell^k\|_1, \quad k \neq \ell, \\ c_\ell(\tilde{\mathbf{X}}^*) &= (c_\ell(\mathbf{X}^*) + \sum_{k \in [m] \setminus \ell} S_k^\ell, (\bar{\mathbf{r}}_\ell)_{n+1}). \end{aligned}$$

Because the $T_{\{k\}}$ are disjoint, we first choose S_k^ℓ satisfying the above equations, and then choose $\{\tilde{X}_v^* : v \in T_{\{k\}}\}$ satisfying the marginals S_k^ℓ . By simple calculation, we have

$$\begin{aligned} \|S_k^\ell\|_1 &= (\mathbf{r}_k)_{n+1} = \frac{1}{m-1} \Sigma_r - \|\mathbf{r}_k\|_1 - \frac{1}{m-1} s; \\ \sum_{\ell=1, \ell \neq k}^m \|S_k^\ell\|_1 &= \|\mathbf{r}_\ell - c_\ell(\mathbf{X}^*)\|_1. \end{aligned}$$

A trivial construction for S_k^ℓ is

$$S_k^\ell = [\mathbf{r}_\ell - c_\ell(\mathbf{X}^*)] \frac{\|S_k^\ell\|_1}{\|\mathbf{r}_\ell - c_\ell(\mathbf{X}^*)\|_1},$$

when $\|\mathbf{r}_\ell - c_\ell(\mathbf{X}^*)\|_1 > 0$; and $S_k^\ell = \mathbf{0}$, when $\|\mathbf{r}_\ell - c_\ell(\mathbf{X}^*)\|_1 = 0$. Given S_k^ℓ , we arbitrarily choose \tilde{X}_v^* for all $v \in T_{\{\ell\}}$ satisfying their marginals equal S_k^ℓ , i.e.,

$$\tilde{X}_v^* = (\mathbf{r}_k)_{n+1} \prod_{\ell \in [m] \setminus k} \frac{S_{k,v_\ell}^\ell}{\|S_k^\ell\|_1}$$

for $v = (v_1, \dots, v_m)$ and $v_k = n+1$, and $(\mathbf{r}_k)_{n+1} > 0$; otherwise $\tilde{X}_v^* = 0$, when $(\mathbf{r}_k)_{n+1} = 0$.

Putting these steps together, we reach the conclusion of the theorem.

B PROOF OF THEOREM 2

First, we introduce some notations that will be used repeatedly in the proof:

$$\begin{aligned} W_S &= \sum_{u \in T_S} \tilde{X}_u^* \quad \forall S \subset [m]; \\ [-i] &= [m] \setminus \{i\} \quad \forall i \in [m]. \end{aligned}$$

Step 1: We prove that for $S, S' \subset [m]$ and $1 \leq |S|, |S'| \leq m-2$, if both W_S and $W_{S'}$ are positive then $S \subset S'$ or $S' \subset S$.

It is equivalent to show that if $W_S, W_{S'} > 0$ then $k := |S \setminus S'| = 0$ or $\ell := |S' \setminus S| = 0$. Assume the contrary that k, ℓ are both positive. Thus, there exist $u \in T_S$ and $v \in T_{S'}$ such that $\tilde{X}_u, \tilde{X}_v > 0$ and using permutation, WLOG we can assume that u and v have the following block forms

$$u = (u_A, u_B, u_C, u_D), \quad v = (v_A, v_B, v_C, v_D),$$

where

$$\begin{aligned}
 u_A &= \underbrace{u_1, \dots, u_i}_{u_j \neq v_j}, & v_A &= (v_1, \dots, v_i), \\
 u_B &= (n+1, \dots, n+1), & v_B &= \underbrace{v_{i+1}, \dots, v_{i+k}}_{\neq n+1}, \\
 u_C &= \underbrace{u_{i+k+1}, \dots, u_{i+k+\ell}}_{\neq n+1}, & v_C &= (n+1, \dots, n+1), \\
 u_D &= u_{i+k+\ell+1}, \dots, u_m, & v_D &= \underbrace{(v_{i+k+\ell+1}, \dots, v_m)}_{u_j = v_j}.
 \end{aligned}$$

Consider u as $\mathbf{0}$ and v as $\mathbf{1}$ in the cube $\{0, 1\}^{i+k+\ell}$. Since $k, \ell > 0$, both Block B and Block C are non-empty. For some $b \in [i+1, i+k]$, we set \tilde{u}^B to be neighbour of u such that $\tilde{u}_b^B = v_b$ and \tilde{v}^B to be neighbour of v such that $\tilde{v}_b^B = u_b = n+1$. For some $c \in [i+k+1, i+k+\ell]$, we construct \tilde{u}^C and \tilde{v}^C similarly. As a result, edges (u, \tilde{u}^B) and (u, \tilde{u}^C) are parallel to edges (v, \tilde{v}^B) and (v, \tilde{v}^C) , respectively. Applying the mass-moving procedure in Lemma 1(a) twice, the total cost is changed by

$$\frac{\epsilon}{2} [D_{k+j-1} + D_{k+j+1} - 2D_{k+j}] + \frac{\epsilon}{2} [D_{\ell+j-1} + D_{\ell+j+1} - 2D_{\ell+j}].$$

which is negative due to the concavity of the sequence $\{D_j\}_{j=1}^{m-1}$. Here the condition $|S|, |S'| \leq m-2$ guarantees that $k+j+1, \ell+j+1 \leq m-1$. Thus the total cost decreases, contradicted to the optimality of $\bar{\mathbf{X}}^*$. Hence, $k=0$ or $\ell=0$, leading to $S' \subset S'$ or $S \subset S'$. Define

$$\begin{aligned}
 \mathcal{T} &= \{S : S \in [m], W_S > 0\}, \\
 \mathcal{T}^- &= \mathcal{T} \setminus \{-1, [-2], \dots, [-m], [m]\}.
 \end{aligned}$$

Note that if $S \in \mathcal{T}^-$, then $|S| \leq m-2$. Due to the inclusion property for any two elements of \mathcal{T}^- , all elements of \mathcal{T}^- could be ordered as a sequence, where the inclusion defines the order in \mathcal{T}^- . Hence, without loss of generality, we could assume that

$$\mathcal{T} \subset \{?, [1], [2], \dots, [m-2], [-1], [-2], \dots, [-m], [m]\}.$$

Step 2: We prove that $W_{[m]} = 0$.

This is equivalent to prove that $\bar{X}_v^* = 0$, where $v = (n+1, \dots, n+1)$. Assume the contrary that $\bar{X}_v^* > 0$. Assume that there exists $u \in T_S$ for $S \in \{?, [1], \dots, [m-2]\}$ such that $\bar{X}_u^* > 0$.

Applying Lemma 1(c) with u as $\mathbf{0}$ and v as $\mathbf{1}$. If $u \in T_?$, then the total cost is changed by

$$-\frac{\epsilon}{m-1} C_u - \epsilon \cdot D_m + \frac{\epsilon \cdot m}{m-1} D_{m-1},$$

which is negative due to $C_u \geq 0$ and $D_m > 0$ and $D_{m-1} = 0$, contradicted to the optimality of $\bar{\mathbf{X}}^*$.

If $u \in T_{[i]}$ for $1 \leq i \leq m-2$, then the total cost is changed by

$$-\frac{\epsilon}{m-i-1} D_i - \epsilon \cdot D_m + \epsilon \frac{m-i}{m-i-1} D_{m-1}.$$

which is negative due to $D_i > 0$, $D_m > 0$ and $D_{m-1} = 0$, contradicted to the optimality of $\bar{\mathbf{X}}^*$.

Hence, either $W_{[m]} = 0$ or $W_S = 0$ for all $S \in \{?, [1], \dots, [m-2]\}$. For the second case, from the equations for the marginals of $\bar{\mathbf{X}}^*$, we have

$$\begin{aligned}
 W_{[-i]} &= \|\mathbf{r}_i\|_1; \quad i = 1, 2, \dots, m \\
 \sum_{j=1}^m W_{[-j]} - W_{[-i]} + W_{[m]} &= \sum_{j=1}^m \|\mathbf{r}_j\|_1 - \|\mathbf{r}_i\|_1 - (m-1)s.
 \end{aligned}$$

Adding all the $W_{[-j]}$, we get $\sum_{j=1}^m W_{[-j]} = \sum_{j=1}^m \|\mathbf{r}_j\|_1$. Comparing with the last equation, we deduce that $W_{[m]} = -(m-1)s$, then $W_{[m]} = 0$ for $s \geq 0$. Overall, in all the cases, we obtain $W_{[m]} = 0$.

Step 3: We prove that $W_{\mathcal{?}} = s$.

Recall that $W_{\mathcal{?}} = \sum_{v \in [n]^m} \bar{X}_v^*$. Note that at the moment, for all $S \subset [m]$, the quantity W_S is zero except for $W_{[i]}$ where $i \in [m-2]$ and $W_{[-i]}$ where $i \in [m]$. Thus, considering only these non-zero terms, we can deploy the system of equations from the marginals as follows:

(1) For the extended marginal $\bar{\mathbf{r}}_1^{(2)}$,

(1.1) for the first n elements, which corresponds to coordinates u satisfying $u_1 \neq n+1$, the equation is

$$W_{\mathcal{?}} + W_{[-1]} = \|\mathbf{r}_1\|_1;$$

(1.2) for the last element, which corresponds to coordinates u satisfying $u_1 = n+1$, the equation is

$$\sum_{i=1}^{m-2} W_{[i]} + \sum_{i=2}^m W_{[-i]} = \sum_{i=2}^m \|\mathbf{r}_i\|_1 - (m-1)s.$$

(2) For the extended marginals $\bar{\mathbf{r}}_i^{(2)}$ (where $i \in \{2, \dots, m\}$), which corresponds to coordinates $u = (u_1, \dots, u_m)$ satisfying $u_i \neq n+1$, the equations are

$$\begin{aligned} W_{\mathcal{?}} + \sum_{j=1}^{i-1} W_{[j]} + W_{[-i]} &= \|\mathbf{r}_i\|_1, \quad 2 \leq i \leq m-2, \\ W_{\mathcal{?}} + \sum_{j=1}^{m-2} W_{[j]} + W_{[-i]} &= \|\mathbf{r}_{m-1}\|_1, \quad i \in \{m-1, m\}. \end{aligned}$$

Taking the sum of the last $m-1$ equations for $i = 2, \dots, m$, we obtain

$$(m-1)W_{\mathcal{?}} + \sum_{i=1}^{m-2} (m-i)W_{[i]} + \sum_{i=2}^m W_{[-i]} = \sum_{i=2}^m \|\mathbf{r}_i\|_1$$

Compare it with the second equation,

$$\sum_{i=1}^{m-2} W_{[i]} + \sum_{i=2}^m W_{[-i]} + (m-1)s = \sum_{i=2}^m \|\mathbf{r}_i\|_1,$$

we obtain that

$$(m-1)W_{\mathcal{?}} + \sum_{i=1}^{m-2} (m-i-1)W_{[i]} = (m-1)s.$$

There are two scenarios:

- $W_{\mathcal{?}} < s$. Since $\|\mathbf{r}_1\|_1 \geq s$, we have $\|\mathbf{r}_1\|_1 > W_{\mathcal{?}}$. Recall that $W_{\mathcal{?}} + W_{[-1]} = \|\mathbf{r}_1\|_1$, hence $W_{[-1]} > 0$. Furthermore, $W_{\mathcal{?}} < s$ implies that there exists $i \in [m-2]$ that $W_{[i]} > 0$.
- $W_{\mathcal{?}} = s$. Since $W_{[i]} \geq 0$ for all $i \in [m-2]$, we have $W_{\mathcal{?}} = s$ and $W_{[i]} = 0$ for all $i \in [m-2]$.

The second scenario is exactly what we want to prove. Thus, we will show that the first scenario is impossible.

For $S = [i]$ and $S' = [-1]$, we consider some $u \in T_S$ as $\mathbf{1}_k$ and some $v \in T_{S'}$ as $\mathbf{0}_k$ in the cube $\{0, 1\}^k$ (where $k = m-i+1$), which have the forms

$$\begin{aligned} u &= (n+1, \underbrace{n+1, \dots, n+1}_{i-1 \text{ times}}, \underbrace{u_{i+1}, \dots, u_m}_{\neq n+1}) \\ v &= (\underbrace{v_1}_{\neq n+1}, \underbrace{n+1, \dots, n+1}_{i-1 \text{ times}}, n+1, \dots, n+1) \end{aligned}$$

Since $W_S, W_{S'} > 0$, there exists $u \in T_S$ and $v \in T_{S'}$ such that $\bar{X}_u^*, \bar{X}_v^* > 0$. Applying the procedure in Lemma 1(c) and noting that $D_{m-1} = 0$, the total cost is changed at least by

$$-\epsilon \cdot D_i + \frac{\epsilon}{k-1} [D_{i-1} + (k-1)D_{i+1}] = \frac{\epsilon}{k-1} [D_{i-1} + (k-1)D_{i+1} - (k-1)D_i].$$

Let $\Delta_j^{(1)} := D_{j+1} - D_j$ for $0 \leq j \leq m-2$ that leads to $\Delta_j^{(2)} := \Delta_j^{(1)} - \Delta_{j-1}^{(1)}$ for $1 \leq j \leq m-2$. Note that $k = m - i + 1$ where $i \in [m-2]$ and $D_{m-1} = 0$, the term inside the above bracket can be written as

$$\begin{aligned} & (m-i)\Delta_i^{(1)} + (D_{i-1} - D_{m-1}) \\ &= (m-i)\Delta_i^{(1)} - [\Delta_{i-1}^{(1)} + \Delta_i^{(1)} + \dots + \Delta_{m-2}^{(1)}] \\ &= \Delta_i^{(2)} - \sum_{j=i+1}^{m-2} [\Delta_{i+1}^{(2)} + \dots + \Delta_j^{(2)}] \\ &= \Delta_i^{(2)} - \sum_{j=i+1}^{m-2} (m-1-j)\Delta_j^{(2)} \\ &= [\Delta_i^{(2)} - (m-1-i)\Delta_{i+1}^{(2)} + \Delta_{i+1}^{(2)}] - \sum_{j=i+2}^{m-2} (m-1-j)\Delta_j^{(2)} \\ &\leq \Delta_{i+1}^{(2)} - \sum_{j=i+2}^{m-2} (m-1-j)\Delta_j^{(2)} \\ &\leq \dots \leq \Delta_{m-2}^{(2)} \leq 0, \end{aligned}$$

where the inequalities in the last two lines come from the condition that $\Delta_j^{(2)} \leq (m-1-j)\Delta_{j+1}^{(2)}$.

Thus, we have proved that

$$\min_{\bar{\mathbf{X}} \in \Pi(\bar{\mathbf{r}}_1, \dots, \bar{\mathbf{r}}_m)} \langle \bar{\mathbf{C}}^{(2)}, \bar{\mathbf{X}} \rangle = \langle \mathbf{C}, \bar{\mathbf{X}}_{[n]^m}^* \rangle \geq \min_{\mathbf{X} \in \Pi^s(\mathbf{r}_1, \dots, \mathbf{r}_m)} \langle \mathbf{C}, \mathbf{X} \rangle.$$

Step 4: We prove the claimed statement. The next step is to prove the inverse inequality in order to deduce the equality.

For a \mathbf{X}^* such that $c_k(\mathbf{X}^*) \leq \mathbf{r}_k$ and $\sum_{v \in [n]^m} X_v^* = s$. Then transport tensor map $\tilde{\mathbf{X}}^*$ could be obtained from expanding the tensor matrix \mathbf{X}^* as follows: For $u = (n+1, \dots, n+1, i_k, n+t, \dots, n+1) \in T_{[-k]}$ and $1 \leq i_k \leq n$

$$\tilde{X}_u^* = (\mathbf{r}_k)_{i_k} - \sum_{v \in [n]^m, v_k = i_k} X_v^*.$$

For other $u \notin [n]^m \cup_{k=1}^m T_{[-k]}$, we set $\tilde{X}_u^* = 0$. Hence, the $\tilde{\mathbf{X}}^*$ satisfies the marginal constraints. It means that

$$\langle \mathbf{C}, \mathbf{X}^* \rangle = \langle \bar{\mathbf{C}}^{(2)}, \tilde{\mathbf{X}}^* \rangle \geq \min_{\mathbf{X} \in \Pi(\bar{\mathbf{r}}_1, \dots, \bar{\mathbf{r}}_m)} \langle \bar{\mathbf{C}}^{(2)}, \mathbf{X} \rangle.$$

Putting these steps together, we obtain the conclusion of the theorem.

C PROOF OF REMAINING RESULTS

C.1 Proof of Lemma 1

First, let F be the face that we take the sum. Note that F consists of vertices in the form (i_1, \dots, i_k) where one of the i_ℓ is constant.

(a) Let F be a face of the cube, WLOG we have

$$F = \{(0, i_2, \dots, i_m)\}$$

Case 1. $\tilde{u} = (0, 1, 0, \dots, 0)$, F loses $\frac{1}{2}\epsilon$ mass at $(0, \dots, 0)$ and gains $\frac{1}{2}\epsilon$ mass at $(0, 1, 0, \dots, 0)$. So the total mass on F are unchanged.

Case 2. $\tilde{u} = (1, 0, \dots, 0)$, F loses $\frac{1}{2}\epsilon$ mass at $(0, \dots, 0)$ and gains $\frac{1}{2}\epsilon$ mass at $(0, 1, \dots, 1)$, since the edge between $(0, \dots, 0)$ and $(1, 0, \dots, 0)$ is parallel to the edge between $(1, \dots, 1)$ and $(0, 1, \dots, 1)$.

(b) Note that if we do it multiple time of moving masses from u and v to theirs neighbours with different $\delta = \frac{1}{k}\epsilon$ mass, the property of unchanged mass sum on a given face still holds.

(c) Due to the symmetric of the transportation for case (c), we could assume that $\ell = 1$. If $i_1 = 0$, then the face F contains all vertices of $(0, i_2, \dots, i_m)$. Apparently, F contains $(0, \dots, 0)$ and only one neighbour of $\mathbf{1}_k$ which is $(0, 1, \dots, 1)$. Hence the quantity $\sum_{v \in F} M_v$ is unchanged.

If $i_1 = 1$, the face F contains all vertices of $(1, i_2, \dots, i_m)$. Here, F contains vertex $\mathbf{1}_k$ and $m - 1$ vertices which are neighbours of $(1, \dots, 1)$, F does not contain $(0, 1, \dots, 1)$ which is also neighbour of $(1, 1, \dots, 1)$. Overall, the total weight M_v of the face F are unchanged, since $-\epsilon + (m - 1)\frac{\epsilon}{m-1} = 0$.

As a consequence, we reach the conclusion of the lemma.

C.2 Proof of Proposition 2

From (Lin et al., 2019a, Theorem 4.5), we have

$$\langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^k \rangle - \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^* \rangle \leq \varepsilon$$

within $O\left(\frac{m^3(n+1)^m \|\mathbf{C}\|_\infty^2 \log(n+1)}{\varepsilon^2}\right)$ arithmetic operations where $\bar{\mathbf{C}}$ is a cost tensor in either Theorem 1 or Theorem 2.

From Theorem 1 and Theorem 2, $\langle \mathbf{C}, \mathbf{X}^* \rangle = \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^* \rangle$. Moreover, since \mathbf{C} is a sub-tensor of the non-negative tensor $\bar{\mathbf{C}}$ by construction, and $\bar{\mathbf{X}}^k$ is also non-negative, we get $\langle \mathbf{C}, \mathbf{X}^k \rangle \leq \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^k \rangle$. Hence, $\langle \mathbf{C}, \mathbf{X}^k \rangle - \langle \mathbf{C}, \mathbf{X}^* \rangle \leq \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^k \rangle - \langle \bar{\mathbf{C}}, \bar{\mathbf{X}}^* \rangle \leq \varepsilon$, completing the proof.